

## 1.1

## Bar Graphs and Dot Plots

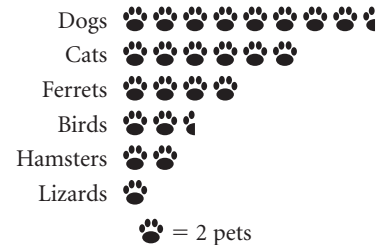
In this lesson you will

- **interpret** and **create** a variety of graphs
- find some **summary values** for a data set
- **draw conclusions** about a data set based on graphs and summary values

This **pictograph** shows the number of pets of various types that were treated at Uptown Animal Hospital in one week.

Specific information, such as the number of pets of each type, is called **data**. You can often display data in tables and graphs.

**Pets Treated in One Week**



**EXAMPLE** | Create a table of data and a **bar graph** from the pictograph above.

► **Solution**

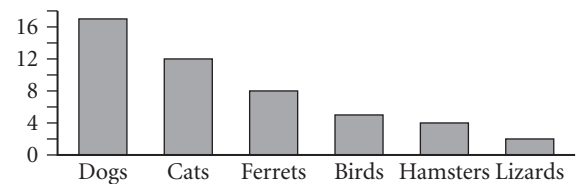
This table shows the number of pets of each type. Remember that each symbol on the pictograph represents two pets.

The bar graph shows the same data. The height of a bar shows the total in that **category**, in this case, a particular type of pet. You can use the *scale* on the *vertical axis* to measure the height of each bar.

**Pets Treated in One Week**

Dogs	Cats	Ferrets	Birds	Hamsters	Lizards
17	12	8	5	4	2

**Pets Treated in One Week**



Bar graphs gather data into categories, allowing you to quickly compare values for each category.

A **dot plot** shows each item of *numerical* data above a *number line*, or *horizontal axis*. Dot plots make it easy to see gaps and clusters in a data set, as well as to see how the data **spreads** along the axis.

### Investigation: Picturing Pulse Rates

This investigation uses data about pulse rates. Pulse rates vary, but a healthy person at rest usually has a pulse rate between certain values. A person with a pulse rate that is too fast or too slow may need medical attention.

(continued)

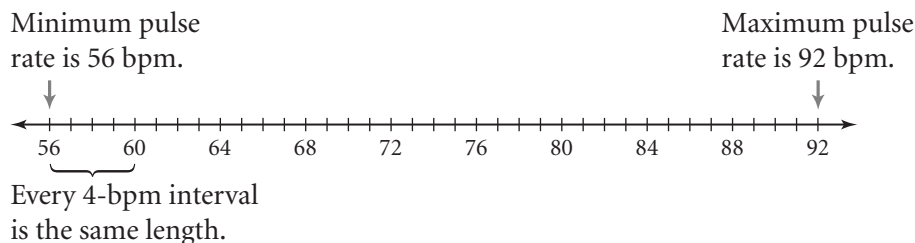
## Lesson 1.1 • Bar Graphs and Dot Plots (continued)

This data set gives pulse rates, in beats per minute (bpm), for a group of 30 students.

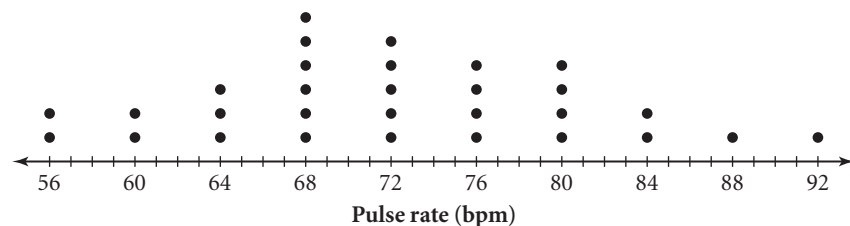
68 60 76 68 64 80 72 76 92 68 56 72 68 60 84  
72 56 88 76 80 68 80 84 64 80 72 64 68 76 72

For this data set, the **minimum** (lowest) value is 56 and the **maximum** (highest) value is 92. The minimum and maximum describe the spread of the data. For example, you could say, “The pulse rates are between 56 and 92 bpm.” Based only on this data, it appears that a pulse rate of 80 bpm would be “normal,” whereas a pulse rate of 36 bpm would be too low.

To make a dot plot of the pulse rates, first draw a number line with the minimum value, 56, at the left end. Select a scale and label equal **intervals** until you reach the maximum value, 92.



For each value in the data set, put a dot above that value on the number line. When a value occurs more than once, stack the dots. For example, the value 64 occurs three times in the data set, so there are three dots above 64. Be sure to label the axis so that it is clear what the data are.



The **range** of a data set is the *difference* between the maximum and minimum values. For this data, the range is  $92 - 56$  or 36 bpm. Note that the range does not tell you about the actual values in a data set. If a paramedic tells you that normal pulse rates have a range of 12, you would not know anything about the minimum or maximum value or any values in-between.

Looking at the dot plot, you can see that there are a few values near the maximum and minimum, but most values cluster between 64 and 80. The value 68 occurs most often, followed by 72. This information gives you an idea of what is a normal pulse for this class, but it could not be used to say what a “normal” pulse is for all people. Many factors, including age, affect pulse rates.

**Statistics** is a word we use many ways. It is often used to refer to numbers that describe or summarize data. For example, you could collect pulse-rate data from thousands of people and then determine a single value that can be considered “normal.” This single value can be called a statistic.

## 1.2

# Summarizing Data with Measures of Center

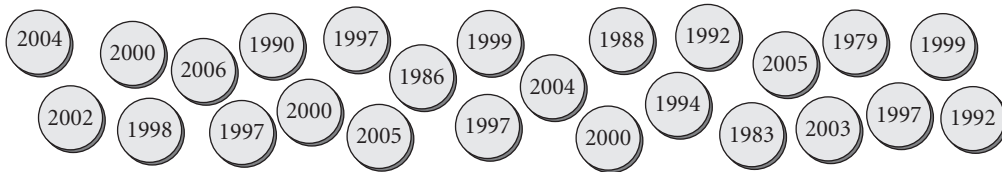
In this lesson you will

- find **measures of center** for a data set
- choose the most meaningful **measure of center** for a situation
- look at the influence of **outliers** on the **mean** of a data set

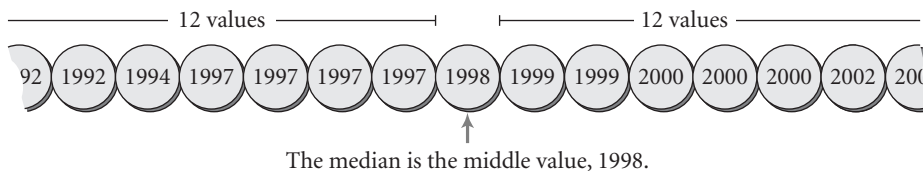
Read the statements at the beginning of the lesson. Each statement uses a single number—called a **measure of center**—to describe what is typical about a set of data. The first statement uses the **mean**, or average. The second statement uses the **median**, or middle value. The third statement uses the **mode**, or most frequent value.

## Investigation: Making “Cents” of the Center

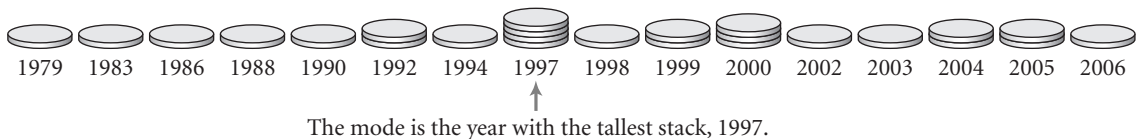
You can use this collection of pennies to help you understand the mean, median, and mode.



To find the median of the mint years, line up the pennies from oldest to newest.



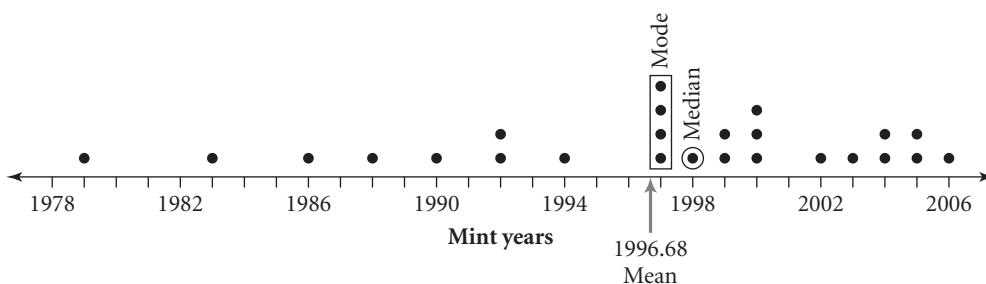
To find the mode, stack the pennies with the same mint year.



To find the mean, find the sum of the mint years for all the pennies and divide by the number of pennies.

$$\text{mean} = \frac{\text{sum of mint years}}{\text{number of pennies}} = \frac{49,917}{25} = 1996.68. \text{ You might round the mean up to 1997.}$$

Here is a dot plot of the mint years with the median, mode, and mean labeled.



(continued)

## Lesson 1.2 • Summarizing Data with Measures of Center (continued)

Now enter the penny data into a calculator list, and use your calculator to find the mean and median. (See **Calculator Notes 1B** and **1C**. Refer to **Calculator Note 1A** to check the settings on your calculator.)

Read the descriptions of the measures of center on page 47 of your book. The mean and the median of a set of data may be quite different, and the mode, if it exists, may not be useful. You will need to decide which measure is most meaningful for a given situation.

### EXAMPLE

This data set shows the number of videos rented at a video store each day for 14 days. {54, 75, 2, 68, 49, 53, 5, 63, 54, 70, 65, 68, 71, 60}

- Find the measures of center.
- Which measure best represents the data?

### ► Solution

- The mean is about 54 videos.

$$\frac{\overbrace{54 + 75 + 2 + 68 + 49 + 53 + 5 + 63 + 54 + 70 + 65 + 68 + 71 + 60}^{\text{Sum of the data values}}}{\underset{\text{Number of data values}}{14}} \approx 54 \quad \text{Mean}$$

Because there are an even number of values, the median is the number halfway between the two middle values. In this case, the median is 61.5.

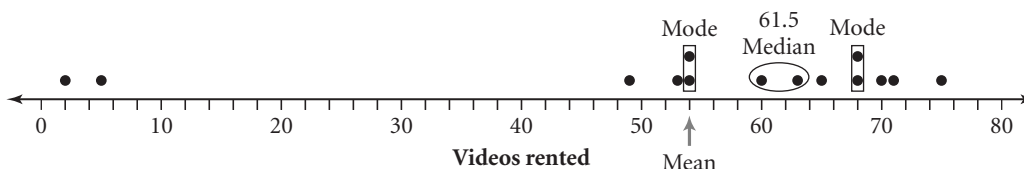
Data values listed in order

2, 5, 49, 53, 54, 54, 60, 63, 65, 68, 68, 70, 71, 75

The median is halfway between 60 and 63.

This data is **bimodal**—that is, it has two modes, 54 and 68.

- To determine which measure best represents the data, look for patterns in the data and consider the shape of the graph.



Except for two values (2 and 5), the data are clustered between 49 and 75. The values 2 and 5, called **outliers**, bring the mean, 54, too far to the left of most of the data values to be the best measure of center. One of the modes is the same as the mean. The other is too high. The median, 61.5, seems to summarize the data best.

In the example in your book, the outliers are much greater than the rest of the data. These values pull the mean up. Using the mean to describe data that includes outliers can be misleading. You need to be aware of these limitations when you read advertisements and reports that give measures of center.

# Five-Number Summaries and Box Plots

In this lesson you will

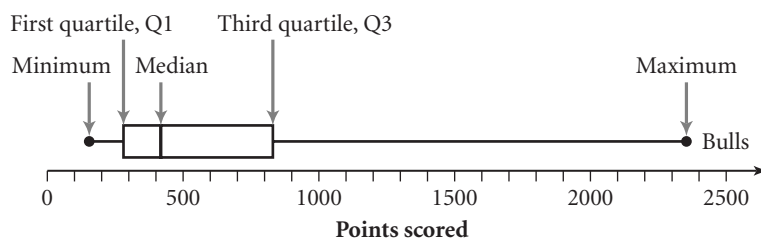
- find **five-number summaries** for data sets
- interpret and create **box plots**
- draw conclusions about a data set based on graphs and summary values
- use a calculator to make a box plot

The table on page 52 of your book shows the total number of points scored by each player on the Chicago Bulls during the 1997–98 NBA season.

The **five-number summary** can give you a good picture of how the Bulls scored as a team. A five-number summary uses five boundary points: the minimum and the maximum, the median (which divides the data in half), the **first quartile** (the median of the first half), and the **third quartile** (the median of the second half).

The first quartile (Q1), the median, and the third quartile (Q3) divide the data into four equal groups. The example in your book illustrates how to find the five-number summary.

A **box plot**, or **box-and-whisker plot**, is a visual way to show the five-number summary. This box plot shows the spread of the Bulls' scoring data. Notice how the five-number summary is shown in the plot.

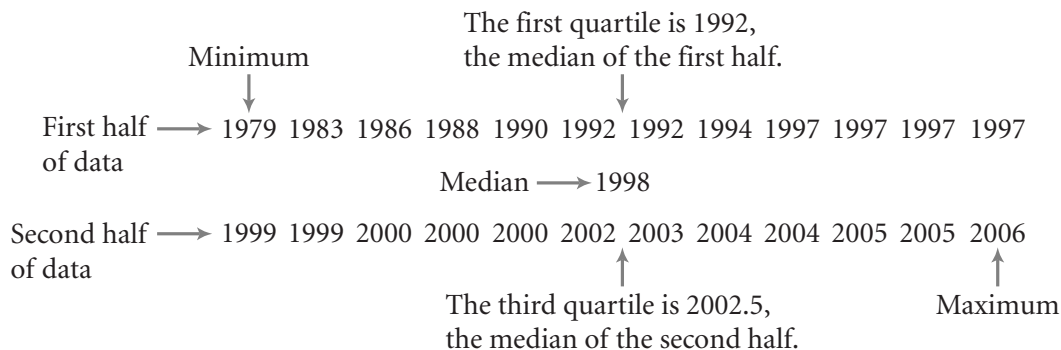


Even though the four sections of the plot—the two whiskers and the two halves of the box—are different lengths, each represents  $\frac{1}{4}$  of the data. So, for example, the long, right whisker represents the same number of data items as the short, left whisker. Notice that most of the data values are concentrated in the left part of the graph. The long whisker that takes up most of the plot represents only  $\frac{1}{4}$  of the values (including Michael Jordan's).

## Investigation: Pennies in a Box

In this investigation, you make a box plot of the penny data from Lesson 1.2.

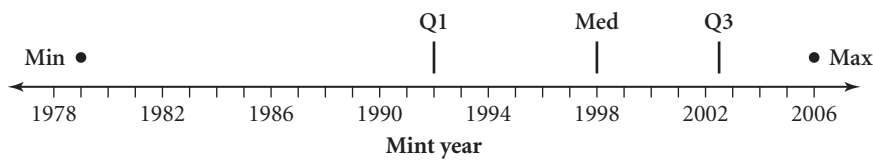
**Steps 1–6** First you need to list the values in order and find the five-number summary. Here is some sample data with the minimum, maximum, median, and quartiles labeled. You may want to follow the steps with your data from Lesson 1.2.



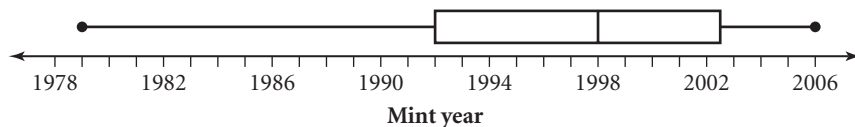
(continued)

## Lesson 1.3 • Five-Number Summaries and Box Plots (continued)

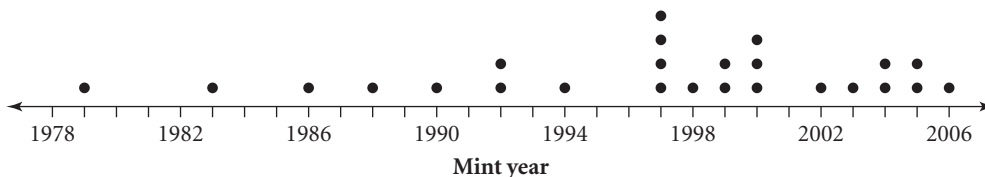
To make a box plot, draw the horizontal axis. (You can use the same scale you used for the dot plot in Lesson 1.2.) Draw a short vertical segment just above the median value. Do the same for the first and third quartiles. Then, draw dots over the minimum and maximum values.



To finish the plot, draw a “box” with ends at the first and third quartiles, and draw “whiskers” extending from the ends of the box to the minimum and maximum values.



Compare the box plot with a dot plot of the same data.



Both plots show that most of the values are greater than 1992. The box plot makes it easier to locate the five-number summary, but, unlike the dot plot, it does not show individual values or how many values there are. If seeing each data value is important, a box plot is not the best way to display your data.

Remember that the four sections of a box plot—the two whiskers and the two halves of the box—each represent about the same number of data items. The box plot shows that  $\frac{3}{4}$  of the mint years are between 1992 and 2006, whereas only  $\frac{1}{4}$  are between 1979 and 1992.

**Steps 7–9** Clear any old data from your calculator and enter the mint years into list L1. Draw a calculator box plot. (See **Calculator Note 1D**.) Compare your calculator plot to the one above. If you use the trace function, you can see the five-number summary for the data.

**Step 10** The difference between the first quartile and the third quartile is the **interquartile range**, or **IQR**. Like the range, the interquartile range helps describe the spread in the data. For the sample penny data above, the range is  $2006 - 1979$  or 27, and the interquartile range is  $2002.5 - 1992$  or 10.5.

You can compare two data sets by making two box plots on the same axis. The box plots on page 54 of your book summarize the final test scores for two algebra classes. You can see that Class A has the greater range of scores and the greater IQR. In both classes, half of the students scored above 80. In Class B, all the students scored above 65, whereas in Class A, fewer than three-quarters of the students scored above 65.

# Histograms and Stem-and-Leaf Plots

In this lesson you will

- interpret and create **histograms**
- interpret **stem-and-leaf plots**
- choose appropriate **bin** widths for histograms
- use a calculator to make histograms

A dot plot includes a dot for each value in a data set. If a data set contains a large number of values, a dot plot might not be practical. A histogram is related to a dot plot but is more useful when you have a large set of data. Read the information about histograms on page 59 of your book. Notice how the bin width chosen for the histogram affects the appearance of the graph.

## Investigation: Hand Spans

Your hand span is the distance from the tip of your thumb to the tip of your little finger when your fingers are spread.

**Steps 1–6** Here are hand spans, to the nearest half-centimeter, of students in one algebra class.

19 18.5 20.5 21.5 18.5 17.5 22 22.5 19.5 20 24 18  
16.5 28 19 20 20.5 24 15 17 19 18 21 21

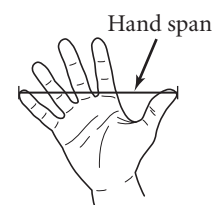
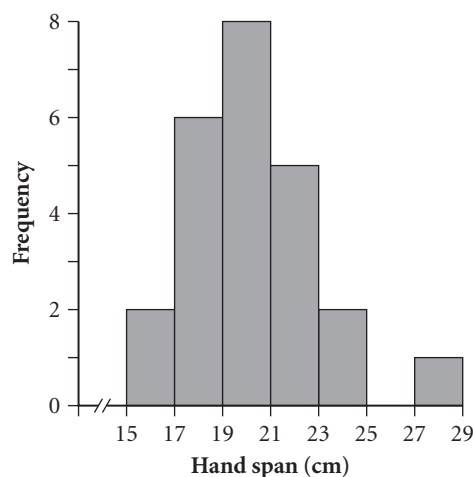
To make a histogram of this data, you first need to choose a bin width. As a general rule, try to choose an interval that will give you from 5 to 10 bins. For this data, the range is  $28 - 15$  or 13. If you use a bin width of 2, you will have 7 bins.

Next, find the number of data values in each bin. These values are called **frequencies**. Note that a bin contains the left boundary value, but not the right. So the bin from 17 to 19 includes the values 17, 17.5, 18, and 18.5, but *not* 19. The value 19 belongs in the bin from 19 to 21.

Bin	15 to 17	17 to 19	19 to 21	21 to 23	23 to 25	25 to 27	27 to 29
Frequency	2	6	8	5	2	0	1

Now, draw the axes. Scale the horizontal axis to show hand-span values from 15 to 29 in intervals of 2. Scale the vertical axis to show frequency values from 0 to 8.

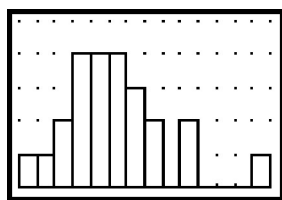
Finally, draw in bars to show the frequency values in your table.



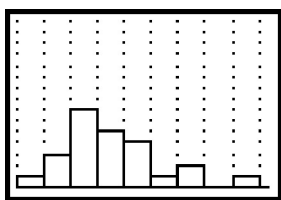
(continued)

## Lesson 1.4 • Histograms and Stem-and-Leaf Plots (continued)

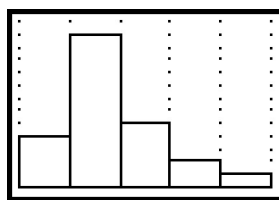
**Steps 7–9** If you enter the hand-span data into list L1 of your calculator, you can experiment with different bin widths. (See **Calculator Note 1E** for instructions on creating histograms.) Here are histograms of the same data using different bin widths.



[15, 29, 1, 0, 5, 1]



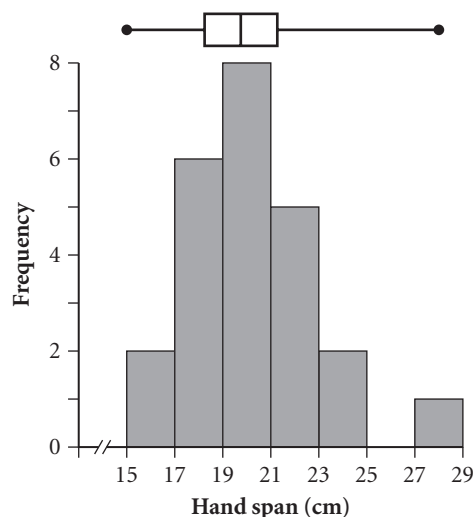
[15, 29, 1.5, 0, 15, 1]



[15, 30, 3, 0, 15, 1]

For this data, a bin width of 1.5 or 2 works well. Both bin widths give a nice picture of how the values are distributed, indicating where data values are clustered and where there is a gap in the values. A bin width of 1 shows additional gaps but has a lot of bars. With a bin width of 3, there are too few bars to get a good picture of the distribution, and the gap on the right side is hidden.

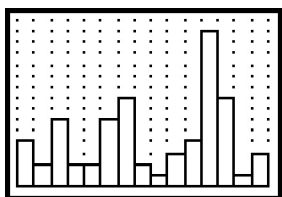
At right, a box plot has been drawn on the same axes as the histogram. Unlike the histogram, the box plot does not show the gap in the data, and it does not give any indication of how many values are in each interval.



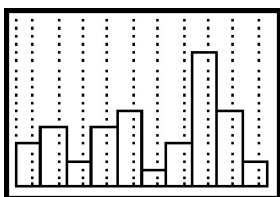
Like a histogram, a **stem plot**, or stem-and-leaf plot, groups data values into intervals. However, a stem plot shows each individual data item. For this reason, stem plots are most useful for fairly small data sets. Study the stem plot on page 61 of your book. Make sure you understand how to read the values.

The example on pages 61–62 of your book shows how to use the data in a stem plot to create a histogram. The values in the stem plot are years. The stem values show the first three digits of the year, and the leaves show the last digit. So, for example, the values next to the stem 185 represent the years 1852, 1852, 1853, and 1857.

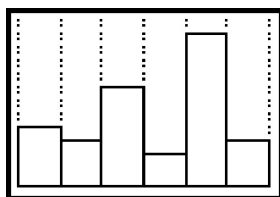
Read through the example to see how the histogram was created. Then, enter the data into your calculator and experiment with different bin widths. Here are some possibilities.



[1850, 2000, 10, 0, 15, 1]



[1850, 2000, 15, 0, 20, 1]



[1850, 2000, 25, 0, 25, 1]

Think about which bin widths best show the spread and distribution of the values.



## 1.6

## Two-Variable Data

In this lesson you will

- practice plotting points on the **coordinate plane**
- create a **scatter plot**
- **look for a relationship** between two variables based on a graph
- use your calculator to create a scatter plot

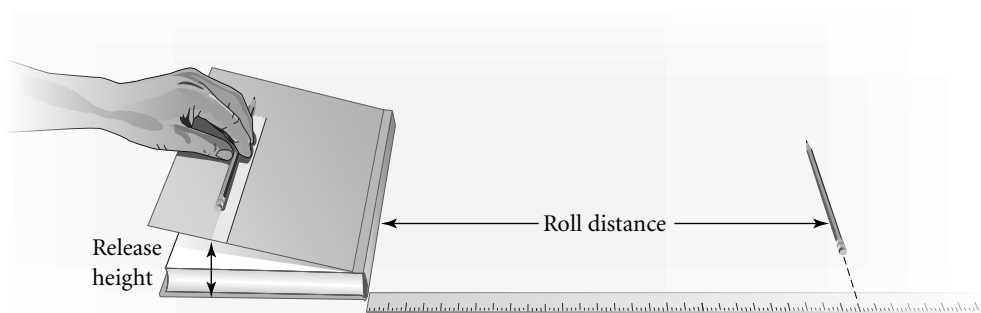
A **variable** is a quantity—such as pulse rate or mint year—whose value can vary. So far in this chapter, you have been working with **one-variable data**. You have used dot plots, box plots, histograms, and stem plots to display one-variable data.

It is often interesting to look at information about *two* variables to try to discover a relationship. For example, you could collect data about height and shoe size to see if taller people tend to have larger feet than shorter people. Making a **scatter plot** is a good way to look for a relationship in **two-variable data**. Read the information about two-variable graphs on page 70 of your book.

### Investigation: Let It Roll!

This investigation involves collecting and graphing two-variable data. Read the introduction to the investigation on page 70 of your book.

**Steps 1–3** To collect the data, mark a point (the “release point”) on the ramp, measure its height, and record that height in a table. Then, let the object roll, measure its stopping distance from the base of the ramp, and record that distance in the table. Repeat with different release points.



Here are some sample data for this activity.

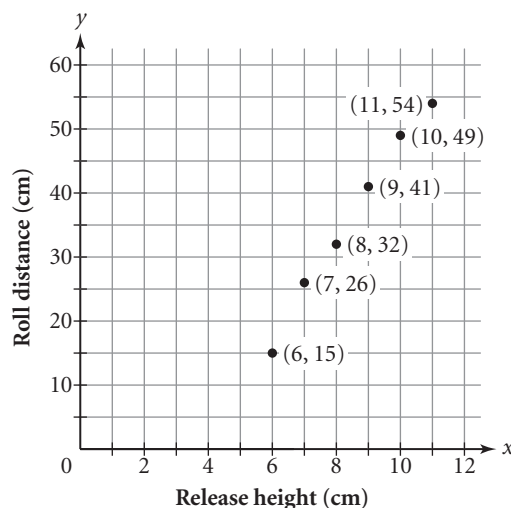
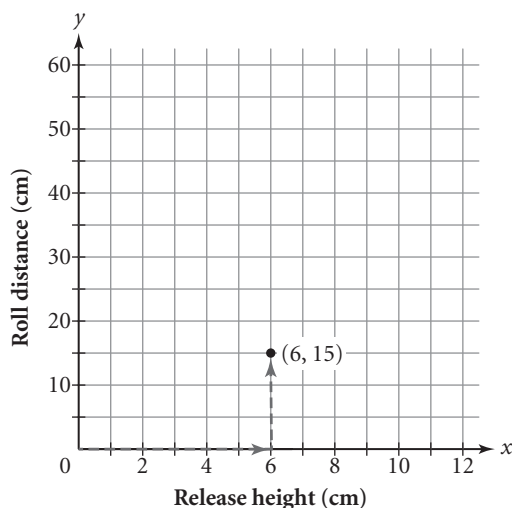
Release height (cm)	6	7	8	9	10	11
Roll distance (cm)	15	26	32	41	49	54

**Steps 4–6** To see if there is a relationship between the variables *release height* and *roll distance*, you can make a scatter plot. You can put either variable on either axis. In the graph on the following page, the release height values are on the  $x$ -axis and the roll distance values are on the  $y$ -axis. The greatest release height value is 11, and the greatest roll distance value is 54. If you let each grid unit on the  $x$ -axis represent 1 cm and each grid unit on the  $y$ -axis represent 5 cm, all the values will fit on a 12-by-12 grid. Be sure to label the axes with the names of the variables.

(continued)

## Lesson 1.6 • Two-Variable Data (continued)

To plot the data, think of each pair of data as an ordered pair. For example, to graph the first pair, plot  $(6, 15)$  by moving along the horizontal axis to 6 and then straight up until you are even with 15. Below on the right is the completed scatter plot.



**Steps 7–8** Notice that the points form an approximate straight-line pattern. The graph indicates that the greater the release height of an object, the greater its roll distance will be.

To make a scatter plot on your calculator, you need to enter the data into two lists. Try making a scatter plot of the data above on your calculator. (See **Calculator Note 1F** to learn how to make a scatter plot.)

Review the example on page 72 of your book. The graph in the example, like the graph created in the investigation, is a **first-quadrant graph** because all the values are positive. Sometimes data includes negative values and requires more than one quadrant.

### EXAMPLE

If you write checks for more money than you have in your account, the account will have a negative balance. The graph at right shows the balance in a checking account at the end of each day over a five-day period. Describe the balance each day.

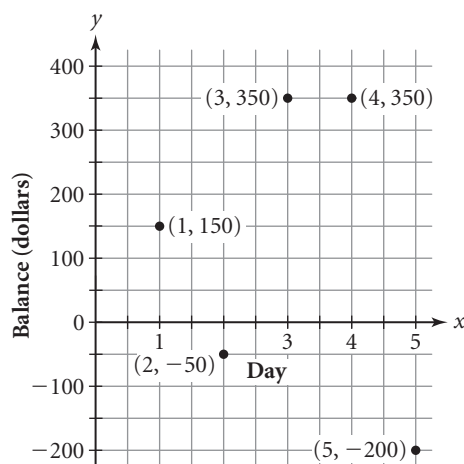
### ► Solution

The point  $(1, 150)$  shows that on the first day, the balance was \$150.

The point  $(2, -50)$  shows that on the second day, the balance was  $-\$50$ .

The points  $(3, 350)$  and  $(4, 350)$  indicate that on the third and fourth days, the balance was \$350.

The point  $(5, -200)$  means that on the fifth day, the balance was  $-\$200$ .



## 1.7

## Estimating

In this lesson you will

- practice your **estimation** skills
- create a **scatter plot**
- **draw conclusions** about a data set based on graphs
- use your calculator to create a scatter plot

In this lesson you see how to use a scatter plot to compare estimated values to actual values.

### Investigation: Guesstimating

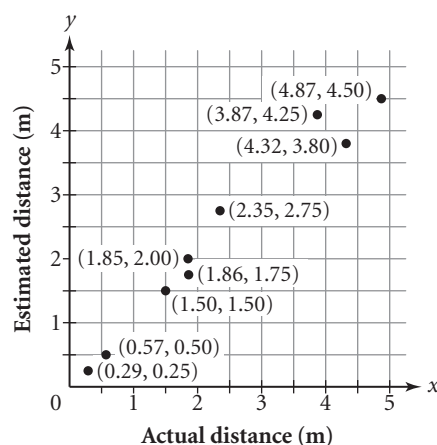
**Steps 1–3** To test your estimation skills, you can estimate the distances from a starting point to several objects and then measure to find the actual distances. This table shows some sample data. You may want to collect and use your own data, or just add a few of your own values to this table.

Description	Actual distance (m), $x$	Estimated distance (m), $y$
Ted's desk	1.86	1.75
Zoe's desk	1.50	1.50
Chalkboard	4.32	3.80
Teacher's desk	1.85	2.00
Blue floor tile	0.57	0.50
Corner of classroom	4.87	4.50
Garbage can	2.35	2.75
Jing's pencil	0.29	0.25
Bookshelf	3.87	4.25

**Steps 4–5** You can make a scatter plot to compare the estimated distances with the actual distances. First, set up the axes. Put the actual distances on the  $x$ -axis and the estimated distances on the  $y$ -axis. Use the *same scale* on both axes.

Plot a point for each pair of values in the table, using the actual distance for the  $x$ -value and the estimated distance for the  $y$ -value. For example, plot (4.32, 3.80) to represent the data for the chalkboard. Here is the completed graph.

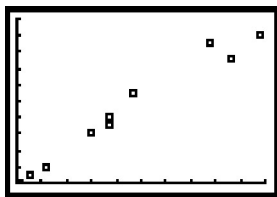
Think about what the graph would look like if *all* of the estimates had been *the same* as the actual measurements. That is, imagine that the graph included (1.86, 1.86), (1.50, 1.50), (4.32, 4.32), and so on. In this case, the points would fall on a straight line.



(continued)

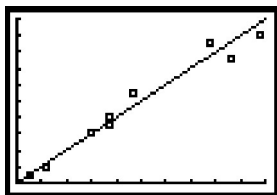
## Lesson 1.7 • Estimating (continued)

**Steps 6–9** Use your graphing calculator to make a scatter plot of the data. Here's what the plot will look like if you use the same scale used in the preceding graph.

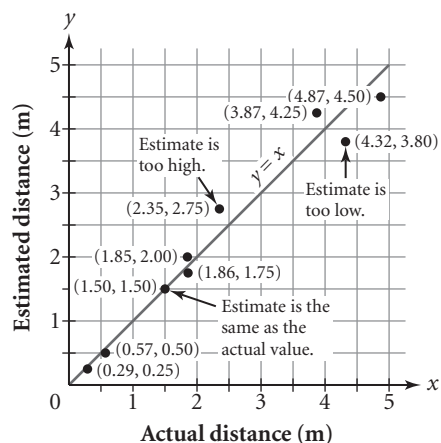


[0, 5, 0.5, 0, 5, 0.5]

The line  $y = x$  represents all the points for which the  $y$ -value equals the  $x$ -value. In this case, it represents all the points for which the estimated distance equals the actual distance. Graph the line  $y = x$  in the same window as your scatter plot. (See **Calculator Note 1J** to graph a scatter plot and an equation simultaneously.)



Points that are closer to the line represent better estimates than points that are farther from the line. Use the trace function to look at the coordinates of the points. Notice that points for which the estimate is too low are below the line and that points for which the estimate is too high are above the line.



# Using Matrices to Organize and Combine Data

In this lesson you will

- use **matrices** to organize and combine data
- use your calculator to **perform matrix calculations**

The table on page 83 of your book shows the average number of hours that people in various countries worked each week in the years 1900 and 2000. Like a table, a **matrix** organizes data into rows and columns. The matrix at right shows the working-hours data.

$$[A] = \begin{bmatrix} 47.7 & 35.9 \\ 56.4 & 30.6 \\ 60.1 & 30.8 \\ 64.2 & 26.9 \\ 55.9 & 37.6 \\ 50.2 & 33.1 \end{bmatrix}$$

This matrix has **dimensions**  $6 \times 2$  because it has six rows and two columns.

Work through Example A in your book. In part c, the working-hours matrix is multiplied by 50. Multiplying a matrix by a number means multiplying each entry by that number. So, to find  $50 \cdot [A]$ , multiply each entry of matrix  $[A]$  by 50. Your calculator can do all the multiplications in one step. (See **Calculator Notes 1L** and **1M**.)

To add or subtract two matrices, add or subtract the corresponding entries. Example B shows how to add two matrices. Read through the example carefully and make sure you understand it. To add or subtract two matrices, the matrices must have the same dimensions. If the dimensions are different, you won't be able to match up corresponding entries to do the operations.

Now you know how to multiply a matrix by a number and how to add or subtract two matrices. As you will see in the investigation, multiplying two matrices is more complicated.

## Investigation: Row-by-Column Matrix Multiplication

**Steps 1–3** The table and matrix show the costs for large items at two restaurants.

Large-Item Prices

	Pizza Palace	Tony's Pizzeria
Large pizza	\$11.40	\$11.35
Large salad	\$ 3.35	\$ 3.90
Large drink	\$ 2.15	\$ 2.10

$$[D] = \begin{bmatrix} 11.40 & 11.35 \\ 3.35 & 3.90 \\ 2.15 & 2.10 \end{bmatrix}$$

The first column of the matrix gives the prices for Pizza Palace. You can find the total cost for 4 large pizzas, 5 large salads, and 10 large drinks by multiplying the number of each item by the cost for each item.

$$\begin{array}{ccccccc} & \text{Pizza} & & \text{Salad} & & \text{Drink} & & \text{Total} \\ & \text{Number} & \text{Cost} & \text{Number} & \text{Cost} & \text{Number} & \text{Cost} & \text{Cost} \\ 4 \cdot 11.40 & + & 5 \cdot 3.35 & + & 10 \cdot 2.15 & = & 83.85 \end{array}$$

(continued)

## Lesson 1.8 • Using Matrices to Organize and Combine Data (continued)

**Steps 4–13** The row matrix  $[A]$  and the column matrix  $[B]$  contain all the information you need to calculate the total food cost at Pizza Palace. Matrix  $[A]$  shows the item quantities, and matrix  $[B]$  shows the costs.

$$[A] = [4 \quad 5 \quad 10] \quad [B] = \begin{bmatrix} 11.40 \\ 3.35 \\ 2.15 \end{bmatrix}$$

Enter  $[A]$  and  $[B]$  into your calculator and find their *product*,  $[A] \cdot [B]$ , or

$$[4 \quad 5 \quad 10] \cdot \begin{bmatrix} 11.40 \\ 3.35 \\ 2.15 \end{bmatrix}$$

(See **Calculator Note 1P** to learn how to multiply two matrices.) You should get the  $1 \times 1$  matrix  $[83.85]$ . The matrix entry, 83.85, is the total price calculated on the previous page. The calculator multiplies each entry in the row matrix by the corresponding entry in the column matrix and adds the results. This is the same calculation.

$$[4 \quad 5 \quad 10] \cdot \begin{bmatrix} 11.40 \\ 3.35 \\ 2.15 \end{bmatrix} = [4 \cdot 11.40 + 5 \cdot 3.35 + 10 \cdot 2.15] = [83.85]$$

You can find the total cost of the food items at Tony's Pizzeria by multiplying the matrices below. Multiply the matrices on your calculator. You should get  $[85.90]$ .

$$[4 \quad 5 \quad 10] \cdot \begin{bmatrix} 11.35 \\ 3.90 \\ 2.10 \end{bmatrix}$$

You can do one matrix calculation to find the total cost for both restaurants.

$$[4 \quad 5 \quad 10] \cdot \begin{bmatrix} 11.40 & 11.35 \\ 3.35 & 3.90 \\ 2.15 & 2.10 \end{bmatrix}$$

Do this calculation on your calculator. You should get the  $2 \times 1$  matrix  $[83.85 \quad 85.90]$ . To get the first entry in the product, the calculator multiplies the row matrix by the first column. To get the second entry, the calculator multiplies the row matrix by the second column. If you try to multiply the matrices in the reverse order, you will get an error message.

To multiply two matrices, the number of columns in the first matrix must be the same as the number of rows in the second because each column entry in the first matrix is multiplied by the corresponding row entry in the second matrix.

Work through Example C in your book to get more practice multiplying matrices.